P1.44 The Issue of Data Density and Frequency with EnKF Radar Data Assimilation in a Compressible Nonhydrostatic NWP Model

Jidong Gao[1] and Ming Xue[1,2]
[1]Center for Analysis and Prediction of Storms and [2]School of Meteorology
University of Oklahoma, Norman, OK 73019

1. Introduction

Since its first introduction by Evensen (1994), the ensemble Kalman filter (EnKF) technique for data assimilation has received much attention. Rather than solving the equation for the time evolution of the probability density function of model state, the EnKF methods apply the Monte Carlo method to estimate the forecast error statistics. A large ensemble of model states are integrated forward in time using the dynamic equations, the moments of the probability density function are then calculated from this ensemble for different times (Evensen 2003).

Recently, EnKF was applied to the assimilation of simulated Doppler radar data for modeled convective storms (Snyder and Zhang 2003; Zhang *et al*. 2004; Tong and Xue 2005, Xue et al. 2006, Gao and Xue 2007) with great successes. But the application to real radar data was not very elegant (Dowell et al. 2004; Tong and Xue 2007). One of the advantages of EnKF method over variational method is that it can dynamically evolve the background error covariances throughout the assimilation cycles, thereby providing valuable uncertainty information on both analysis and forecast. Recently, Caya et al. (2005) showed that with simulated radar data, the EnKF method can outperform a similarly configured 4DVAR scheme after the first few assimilation cycles. When combined with an existing ensemble forecast system (operational ensemble forecasting system is usually run at a lower resolution compared to the operational deterministic forecast), the EnKF method can provide quality analyses with a relatively small incremental cost compared to a 4DVAR system that requires repeated integrations of the forward prediction model and its adjoint.

Same as 4DVAR, the overall computational cost of ensemble-based assimilation methods is significant because of the need for running an ensemble of forecast and analysis of nontrivial sizes (usually a few tens to a few hundreds), especially when high-density data are involved and when the ensemble of all forecasts is run at high resolutions. One of the major sources of errors with the EnKF is the sampling error associated with the limited ensemble size. A larger ensemble helps improve the background error covariance estimation, but incurs high computational cost. To reduce the computational cost and sampling errors, Gao and Xue (2007) proposed a dual-resolution (DR) hybrid ensemble DA strategy that is in a way analogous to the incremental 4DVAR approach. With this strategy, an ensemble of forecasts is run at a lower resolution which provides the background error covariance estimation for both an ensemble of LR analyses and a single HR analysis. On the meso- and convective scales, especially for intense buoyant convection that is both highly nonhydrostatic and intermittent, the WSR-88D Doppler Radar is the only operational instrument capable of providing high spatial and temporal resolution observations, but only radial velocity and reflectivity. The EnKF methods can provide the cross-covariances among the state variables, which can be used to retrieve the variables not directly observed by the radar.

Another big challenge is that the WSR-88D radar network provides a huge amount of data every several minutes across the country. To successfully assimilate these data into numerical models, data thinning can not be avoided. In this work, we study the impact of radar data densities when assimilating them using the ensemble Kalman filter method through OSSEs with the Advanced Regional Prediction System (ARPS). The radar data thinning problem is investigated for the cases where the data density is higher than, lower than or similar to model resolution. The problem of data frequency will be examined in the future.

The rest of this paper is organized as follows. In section 2, we describe the design of the OSS (Observing System Simulation) experiments. Some preliminary experimental results are presented in section 3. Summary and conclusions are given in section 4.

2. Assimilation System

The key to the ensemble-based filter algorithms is the estimation of the background error covariance and the calculation of the Kalman gain matrix using a forecast ensemble. It was first proposed by Evensen (1994). Since then, there have been a number of further developments with the algorithm to ensure that the technique works when the ensemble size is relatively small (Burgers *et al.* 1998; Houtekamer and Mitchell 1998). Whitaker and Hamill (2002) proposed an ensemble square-root filter algorithm (EnSRF) that does not require the perturbation of observations; the assumption that the observational errors are uncorrelated enables the

Corresponding author address: Dr. Jidong Gao, CAPS, Suite 2500, NWC, 120 David L. Boren Blvd, Norman OK 73072. jdgao@ou.edu

processing of the observations serially, one at a time, leading to a considerable simplification of the analysis scheme.

As with all Kalman filter algorithms, the EnSRF algorithm proceeds in two steps, an analysis step and a forecast or propagation step. In the analysis step, the following equations are used to update the state vectors for the ensemble mean and individual ensemble members:

$$\overline{\mathbf{x}}^a = \overline{\mathbf{x}}^b + \mathbf{K}\left[\mathbf{y}^o - H\left(\overline{\mathbf{x}}^f\right)\right], \tag{3}$$

$$\mathbf{x}_n^a = \overline{\mathbf{x}}^a + \beta(\mathbf{I} - \alpha\mathbf{K}\mathbf{H})(\mathbf{x}_n^f - \overline{\mathbf{x}}^f), \tag{4}$$

Here, $\mathbf{x}$ is the state vector we seek to analyze or estimate, and superscripts a and b refer to the analysis (posteriori estimate) and background forecast (prior estimate), respectively, and $\mathbf{y}^o$ is the observation vector, following the standard notation of Ide et al (1997). H is the forward observation operator that maps the model state to the observations, and $\mathbf{H}$ is the linearized version of H. n represents the nth ensemble member and the overbar denotes the ensemble mean. β is a covariance inflation factor that is usually slightly larger than 1. The α is given in the EnSRF algorithm by (Whitaker and Hamill 2002)

$$\alpha = \left[1 + \sqrt{\mathbf{R}(\mathbf{H}\mathbf{P}^b\mathbf{H}^T + \mathbf{R})^{-1}}\right]^{-1}. \tag{5}$$

This procedure produces an ensemble of analyses, as given by Eq. (4). $\mathbf{R}$ and $\mathbf{P}$ are, respectively, the covariance matrices for the observation and background errors. In the forecast step, forecasts are made from each ensemble analysis and are used as the prior estimate or background in the next analysis-forecast cycle; the algorithm continues as the analysis cycles are repeated.

3. Experiment Design and Preliminary Results

In this study, we test the impact of data density to the EnSRF data assimilation using simulated data from a classic May 20, 1977 Del City, Oklahoma supercell storm (Ray *et al.* 1981). Such simulation experiments are commonly referred to as Observing System Simulation Experiments (OSSE, see, e.g., Lord *et al.* 1997). The prediction model, the ARPS, is used in a 3D cloud model mode and the prognostic variables include three velocity components u, v, w, potential temperature θ, pressure p, and six categories of water substances, *i.e.*, water vapor specific humidity q_v, and mixing ratios for cloud water q_c, rainwater q_r, cloud ice q_i, snow q_s and hail q_h.

The truth simulation or nature run is created using the 1 km horizontal resolution, all the model parameter settings and configurations are same as Gao and Xue (2007). As in Snyder and Zhang (2003) and Tong and Xue (2005), the simulated radial velocity, V_r, are assumed to be available on the grid points. Random errors drawn from a normal distribution with zero mean and a standard deviation of 1 ms^{-1} are added to the simulated data. Since V_r is sampled directly from velocity fields, the effect of hydrometeor sedimentation is not involved. The ground-based radar is located at the southwest corner of the computational domain, *i.e.*, at the origin of *x*-*y* coordinate. Different from most existing 4DVAR and EnKF studies, the prediction model is not assumed perfect, the truth simulation is done on 1 km grid, while all EnKF experiments are done on 4 km model grid, so model error caused by different model resolution is quite large.

As an initial effort, we perform three sets of experiments, LR_1km, LR_4km and LR_8km using simulated data every 1 km, 4 km and 8 km respectively in the horizontal. But vertically the simulated data are available every 500 m for all experiments. Also, the data are assumed to be available where reflectivity is greater than 0 dBZ. We start the initial ensemble analysis at 30 min of the model integration time when the storm cell reaches peak intensity.

Figures 1(d, e, f) and 2(d, e, f) show the analyzed fields at the surface and at the 3.5 km level respectively for SR_1km with high density data (1km) being used. it can be seen from Fig. 1d that after 2 analysis cycles, only velocity structure around the storm is somewhat captured while other perturbation fields, such as the perturbation potential temperature and reflectivity, are very weak. The updraft is established pretty well at the 3.5 km level (Fig. 2d) but its structure is smooth (Fig. 2a). After six more analysis cycles, at t = 70 minutes, the low-level flows immediately underneath the storm cells become more stronger (Fig. 1e vs Fig. 1b) although the outflow to the southwest of the storms is generally too weak, and directions of wind vectors more tend to west. Similarly, the extent of the cold pool is too small on the southwest side. However, the difference from the truth is that the left moving storm looks much stronger than the right mover. At the later times, the low-level reflectivity starts to gain a hook echo pattern (Figs. 1c and 1f). Because the EnSRF experiment is performed with 4km model grid, the overall fields are much smoother and the cold pool is too warm even with this high density data.

When the horizontal data density is reduced by a factor of 4 in LR_4km, same resolution as EnSRF analysis, the analyzed low-level cold pool, gust front, and precipitation pattern at the end of the assimilation window become a little more smooth from those of the truth (Fig. 1i); the reflectivity core becomes broader and the hook echo is less well defined. The mid-level updrafts appear only a bit weak, two mesocyclones still look very clear (Figs. 1g-i vs Figs. 1a-b). When we further reduce the data density to 8 km resolution, the analyzed two storms are still there, but much smoother and

weaker (Fig 1j-l and Fig 2j-l). The more thinner of radar data, the more weaker of updraft. However, in the above three experiments, the maximum of reflectivity has not changed too much, and the development of mesocyclone indicated in Fig 2 in all cases remain very clear. Further reduce the data density caused the missing of storms in the analysis. The apparently larger sensitivity of the *w* analysis to the data density appears to be related to the accuracy of horizontal divergence analysis when v_r resolution is significantly reduced. Overall, the horizontal data density of 8 km appears too low for our analysis that uses a 4 km model grid. A 4 km horizontal data resolution is a reasonable compromise between accuracy and cost.

4. Summary and discussion

To successfully assimilate radar data into numerical models, data thinning can not be avoided. In this work, we study the impact of radar data density when assimilating them using the ensemble Kalman filter method through OSSEs with the Advanced Regional Prediction System (ARPS). Since the truth run is from 1 km grid, and EnSRF experiments are performed in 4 km grid, the results are generally smoother than the truth simulation because of model errors. However, the evolution of storm still can be reproduced, especially the low-level cold pool, precipitation (indicated by reflectivity). It is concluded that for storm scale data assimilation, the data density should be at least similar to model resolution. Using the data in much lower density than model resolution may significantly undermine the strength and morphology of the storm. Further experiments will be done with different thinning strategies and different model resolutions and data frequency in the near future.

Acknowledgement

The authors were supported by NSF grants ATM-0331756, ATM-0331594, ATM-0530814 and EEC-0313747, a DOT-FAA grant via DOC-NOAA NA17RJ1227.

References

Burgers, G., P. J. v. Leeuwen, and G. Evensen, 1998: Analysis scheme in the ensemble Kalman filter. *Mon. Wea. Rev.*, **126,** 1719-1724.

Caya, A., J. Sun, and C. Snyder, 2005: A comparison between the 4D-VAR and the ensemble Kalman filter techniques for radar data assimilation. *Mon. Wea. Rev.***, 133**, 3081-3094.

Dowell, D., F. Zhang, L. J. Wicker, C. Snyder, and N. A. Crook, 2004: Wind and thermodynamic retrievals in the 17 May 1981 Arcadia, Oklahoma supercell: Ensemble Kalman filter experiments. *Mon. Wea. Rev.***, 132**, 1982-2005.

Evensen, G., 1994: Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics. *J. Geophys. Res.*, **99(C5),** 10143-10162.

——, 2003: The ensemble Kalman filter: Theoretical formulation and practical implementation. *Ocean Dynamics*, **53,** 343-367.

Gao, J. and M. Xue, 2007: An efficient dual-resolution approach for ensemble data assimilation and tests with assimilated Doppler radar data. *Mon. Wea. Rev.*, Conditionally accepted.

Houtekamer, P. L. and H. L. Mitchell, 1998: Data assimilation using an ensemble Kalman filter technique. *Mon. Wea. Rev.*, **126,** 796-811.

Ide, K., P. Courtier, M. Ghil, and A. Lorenc, 1997: Unified notation for data assimilation: Operational, sequential and variational. *J. Meteor. Soc. Japan*, **75,** 181-189.

Lord, S. J., E. Kalnay, R. Daley, G. D. Emmitt, and R. Atlas, 1997: Using OSSEs in the design of the future generation of integrated observing systems. *Preprint volume, 1st Symposium on Integrated Observation Systems*, Long Beach, CA, Amer. Meteor. Soc.

Hamill, T. M., and C. Snyder, 2000: A hybrid ensemble Kalman filter/3D-variational analysis scheme. *Mon. Wea. Rev.*, **128**, 2905–2919.

Ray, P. S., B. Johnson, K. W. Johnson, J. S. Bradberry, J. J. Stephens, K. K. Wagner, R. B. Wilhelmson, and J. B. Klemp, 1981: The morphology of severe tornadic storms on 20 May 1977. *J. Atmos. Sci.*, **38,** 1643-1663.

Snyder, C. and F. Zhang, 2003: Assimilation of simulated Doppler radar observations with an ensemble Kalman filter. *Mon. Wea. Rev.*, **131,** 1663-1677.

Tong, M. and M. Xue, 2005: Ensemble Kalman filter assimilation of Doppler radar data with a compressible nonhydrostatic model: OSSE Experiments. *Mon. Wea. Rev.*, **133**, 1789-1807.

Whitaker, J. S. and T. M. Hamill, 2002: Ensemble data assimilation without perturbed observations. *Mon. Wea. Rev.*, **130,** 1913-1924.

Xue, M., M. Tong, and K. K. Droegemeier, 2005: An OSSE framework based on the ensemble square-root Kalman filter for evaluating impact of data from radar networks on thunderstorm analysis and forecast. *J. Atmos. Ocean Tech.***, 23**, 46-66.

Zhang, F., C. Snyder, and J. Sun, 2004: Impacts of initial estimate and observations on the convective-scale data assimilation with an ensemble Kalman filter. *Mon. Wea. Rev.*, **132,** 1238-1253.

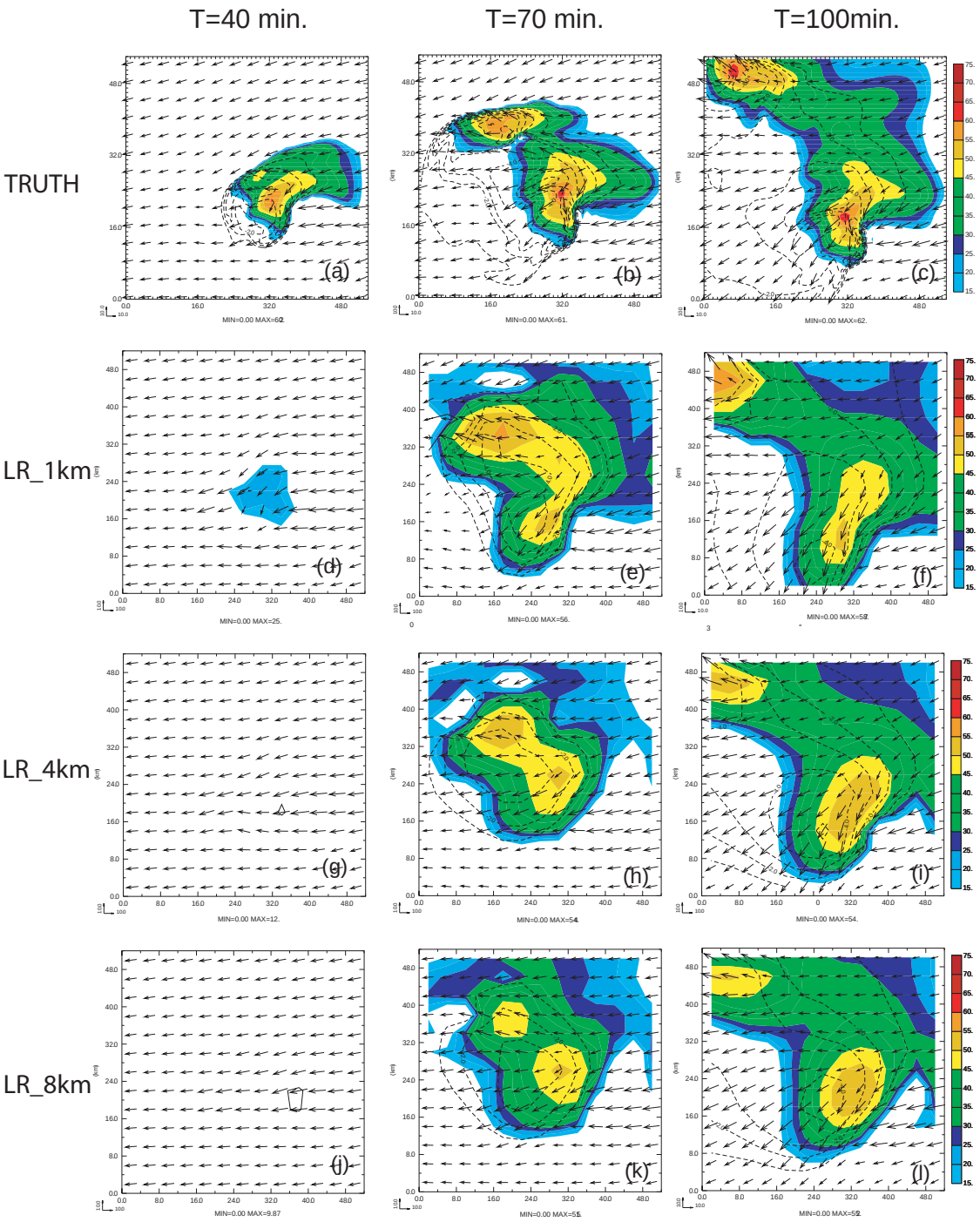

Fig. 1. Horizontal winds (vectors; ms^{-1}), perturbation potential temperature (contours at 1 K intervals) and simulated reflectivity (shaded contours, in dBZ) at 250 m AGL, for the truth simulation (1km) (a)-(c); low-resolution ensemble mean analyses for LR_1km (d)-(f), LR_4km (g)-(i), and SR_8km (j)-(l). The times shown are 40, 70, and 100 min. Wind vectors are shown every 4 km.

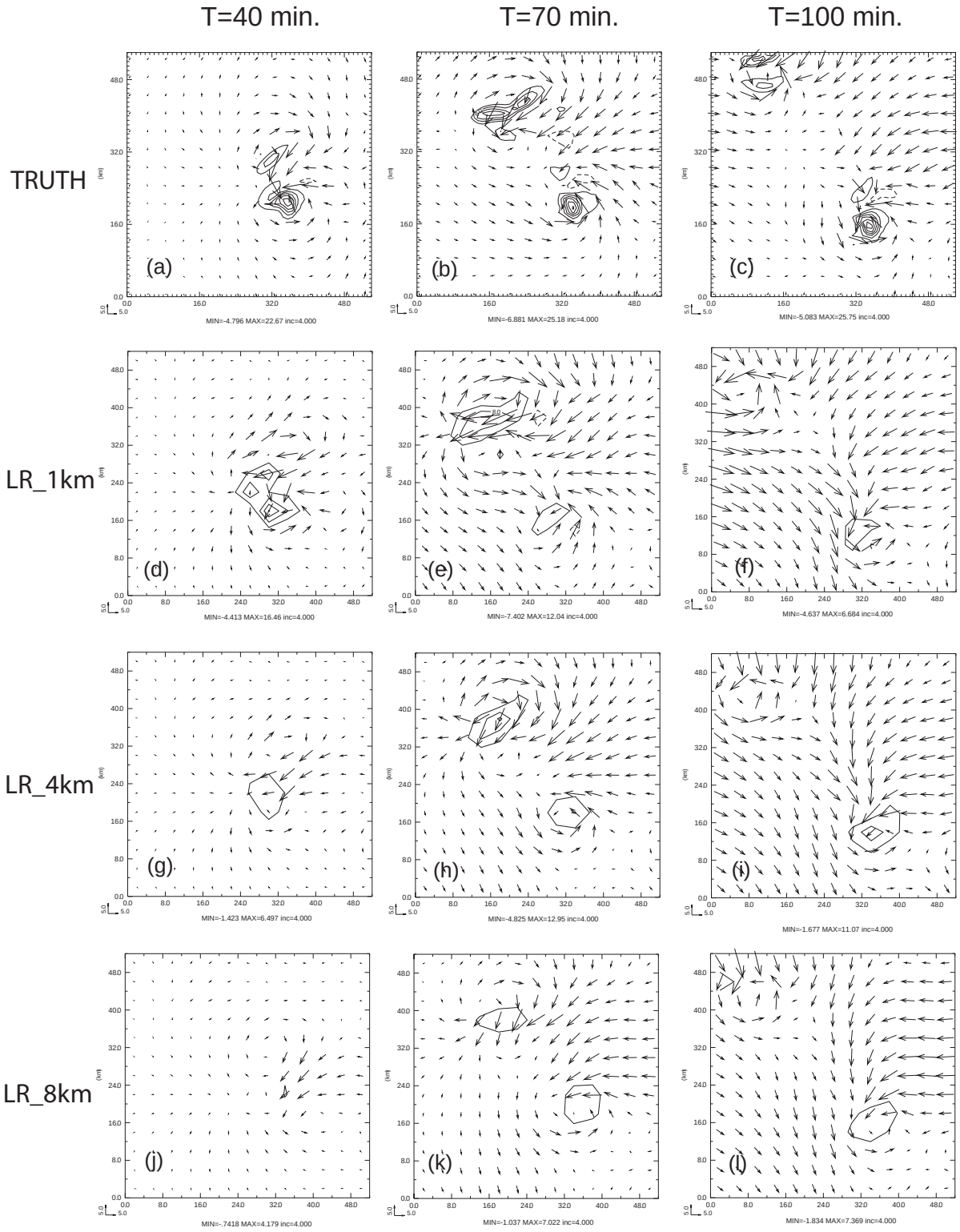

Fig. 2. Same as Fig. 1, but for perturbation horizontal winds from the environment and the vertical velocity w *fields at 3.5 km AGL. The* w *contour interval is 4 ms^{-1} and positive contours are solid and negative contours are dashed.*